A sensitive data access model in support of learning health systems

Ecarot, Thibaud; Fraikin, Benoît; Lavoie, Luc; McGilchrist, Mark; Ethier, Jean François

DOI:
10.3390/computers10030025

Publication date:
2021

computers

MDPI

Article

A Sensitive Data Access Model in Support of Learning Health Systems

Thibaud Ecarot [1], Benoît Fraikin [1], Luc Lavoie [1], Mark McGilchrist [1,2] and Jean-François Ethier [1,*]

[1] Centre Interdisciplinaire de Recherche en Informatique de la Santé, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada; thibaud.ecarot@usherbrooke.ca (T.E.); benoit.fraikin@usherbrooke.ca (B.F.); luc.lavoie@usherbrooke.ca (L.L.); m.m.mcgilchrist@dundee.ac.uk (M.M.)

[2] Department of Health and Clinical Services, University of Dundee, Dundee DD1 4HN, UK

* Correspondence: jean-francois.ethier@usherbrooke.ca

Abstract: Given the ever-growing body of knowledge, healthcare improvement hinges more than ever on efficient knowledge transfer to clinicians and patients. Promoted initially by the Institute of Medicine, the Learning Health System (LHS) framework emerged in the early 2000s. It places focus on learning cycles where care delivery is tightly coupled with research activities, which in turn is closely tied to knowledge transfer, ultimately injecting solid improvements into medical practice. Sensitive health data access across multiple organisations is therefore paramount to support LHSs. While the LHS vision is well established, security requirements to support them are not. Health data exchange approaches have been implemented (e.g., HL7 FHIR) or proposed (e.g., blockchain-based methods), but none cover the entire LHS requirement spectrum. To address this, the Sensitive Data Access Model (SDAM) is proposed. Using a representation of agents and processes of data access systems, specific security requirements are presented and the SDAM layer architecture is described, with an emphasis on its mix-network dynamic topology approach. A clinical application benefiting from the model is subsequently presented and an analysis evaluates the security properties and vulnerability mitigation strategies offered by a protocol suite following SDAM and in parallel, by FHIR.

Keywords: healthcare; protocols; network security; communication system security; data security

Citation: Ecarot, T.; Fraikin, B.; Lavoie, L.; McGilchrist, M.; Ethier, J.-F. A Sensitive Data Access Model in Support of Learning Health Systems. *Computers* **2021**, *10*, 25. https://doi.org/10.3390/computers10030025

Academic Editors: Antonio Celesti, Ivanoe De Falco, Antonino Galletta and Giovanna Sannino

Received: 26 January 2021
Accepted: 23 February 2021
Published: 26 February 2021

1. Introduction

Despite extraordinary research successes, many patients receive sub-optimal care. This includes cases where required action is evident, like for peripheral vascular disease patients and cholesterol lowering drugs (example in [1]) even when there are publications in high "impact" journals. There is, therefore, an urgent need for integrated knowledge transfer tools, like decision support systems and audit-feedback tools, to maximally increase care quality for patients [2]. It is to fill this gap that the Learning Health System (LHS) framework has been developed [3]. In a LHS, the focus is not on a research protocol, but rather on a learning cycle. The cycle workflow starts by looking at data naturally produced during care delivery (increasing pertinence), includes the research activities, and structures knowledge transfer actions right from the planning stage.

Obviously, learning cycles imply intense usages of sensitive health data. The core data requirement, for care delivery, research, and knowledge transfer, is to understand, as best as possible, an individual through their health data. Health is understood here in the larger sense, including data produced by care organisations, as well as research datasets (e.g., biobanks, cohorts), and quantified-self data (e.g., exercise watches, scales, and glucose monitoring devices). All these data are not hosted in a single organisation nor a single system when deploying a LHS at scale.

While many challenges exist regarding data access (like obtaining consent or annotating datasets), communication protocols are critical and the basis on which information

can be exchanged. Making sure that the right protocols are used, and that they fully support LHS, is a critical step when deploying LHSs. To ensure that the correct protocols are used, requirements pertaining to operations and communications' security – in the context of LHSs – need to be modelled and organised. This modelling will enable the sound integration of security protocols. In this article, we present the Sensitive Data Access Model (SDAM) to achieve this goal.

Access to health data across different organisations must abide by different rules emanating from different spheres: Social [4,5], legal [6], or technological [7,8]. SDAM needs to take these into account with a focus on protecting sensitive data. Given the high-level goal of understanding an individual through health data that is hosted in various organisations and the privacy regulations from those spheres, high-level requirements can be proposed with regards to LHSs support.

First, the most efficient way to avoid communication confidentially breach for sensitive data is not to exchange it in the first place. To minimise the quantity of data exchanged, the protocol suite must enable operations to constrain a cohort—based on individuals' characteristics—as precisely as possible. In addition, it must enable data extraction methods that produce the minimal result set required to answer the question at hand. Given current frameworks (both legal and ethical), many projects will require that extracted data be transferred to a well identified organisation that becomes responsible for the received data.

Second, given the highly confidential nature of the data exchanged, the protocol suite should be as resistant as possible to related security threats (e.g., man-in-the-middle attacks). Message payloads should only be available in clear to the originating and final receiving entities, not to intermediate nodes (end-to-end) and the integrity of the message should be guaranteed.

Third, the protocol must be able to be executed purely with synthetic, project-specific identifiers to support a fully deidentified execution. The payload could contain personally identifying data (full name, insurance number, etc.) if the project requires it, but, as mentioned above, this information should be visible only to the message's original sender and final receiver. This information cannot be used for routing or orchestrating activities for example.

At the same time, it is essential that the resulting final dataset provides links across data blocks, coming from various sources, to allow the final receiver to know what data blocks relate to a given individual (data linkage). It should be noted that this data linkage does not imply that the final receiver is allowed to know the identity of an individual, nor from which source a block of data originates from, only that it relates to the same individual. To achieve this, data sources are not allowed, in many circumstances, to know if an individual has data in another source. More generally, different entities contributing or receiving health data should not share an identifier set for the participants.

Fourth, given the scope of the activities covered and reluctance of organisations to deploy ad hoc security technologies (for good reasons), existing security and privacy standards and regulations must be leveraged as much as possible (e.g., those of the Internet Engineering Task Force—IETF).

These high-level requirements imply an important topological requirement. The system should be able to execute a plan without requiring the publication of a data source's IP address. This supports the third requirement of a fully deidentified execution while also minimising the risks of network topology reconstruction which could subsequently enable traffic analysis and man in the middle attacks. This restriction was indeed encountered empirically during the TRANSFoRm project [9] where data sources (primary care electronic medical record systems providers) were reluctant or strictly refused to share their IP address and insisted, instead, that traffic should be routed through a trusted proxy server.

The article is structured as follows. First, published approaches supporting health data access were surveyed to evaluate if one was fulfilling all high-level requirements and could serve as a basis for the development of SDAM. We reviewed both largely implemented approaches, like HL7 standards, as well as more recent technics, like those based on

blockchains. In order to be able to express the specific communication requirements of SDAM, a Data Access System Framework to support LHSs is presented, followed by the specific communication security requirements. Second, the SDAM, based on the Open System Interconnection model (OSI) [10], is subsequently presented: Layers are described, examples of candidate protocols are mentioned, and a focus is placed on the functions and properties of the proxy pool, particularly in the context of data-related activities, like extractions and transmissions, via the clinical application of the model using Reflex-D, an audit-feedback tool addressing diabetes care. Third, a discussion is presented. The discussion starts with a security vulnerability analysis for SDAM. This is followed by an evaluation of the Fast Healthcare Interoperability Resources (FHIR) security properties and vulnerabilities. FHIR was selected given its expanding use in North America and other jurisdictions. The security properties and vulnerabilities selected for presentation are those identified as important and addressable by SDAM. Finally, a summary of the work presented here is offered with future developments.

2. Related Works

Information exchange among health organisations has been studied in various contexts. This review starts with the description of existing approaches currently in use to support health data exchanges in different contexts. Other more recent methods are also presented, but they have been mostly used in contexts of proofs of concept, like blockchains or onion routing networks.

Many data exchange systems are based on existing protocols such as FTP on SSL [11,12]. Other solutions call upon the Secure Shell protocol (SSH), which may turn into a serious issue when used through a virtual private network. This is true in many broad-range healthcare research and care delivery organisations [13]. This particular type of architecture with FTP or SSH over VPN generates risks of man-in-the-middle attacks. A list of threats and potential attacks on the Transport Layer Security (TLS) and Secure Shell protocols is available [14]. However, while Virtual Private Networks (VPNs) usually call upon the L2TP protocol, reference [15] this alone does not suffice to meet with the LHSs' requirements: There are threats like traffic fingerprinting attacks, possible unencrypted intermediate texts (metadata added during transit of network packets), and potentially leaked client identity keys allowing the identification of the sender for a given message [16].

Several health organisations, particularly (but not exclusively) in North America, base their medical data exchanges on the standards developed by HL7. Health Level 7 (HL7) is named after the Application Layer 7 as used in the OSI model. The HL7v2 and HLv3 standards are based on an architectural paradigm focused on message-oriented functionality. These old versions of the standards were defined before the TLS standard and a safety study was carried out. Findings of the TLS safety study found that there are very high security risks present when holding on to these standards in production [17]. For example, the authors showed that it was relatively simple to set up a "man-in-the-middle" TCP attack on HL7v2 exchanges and to undermine the integrity of the data by altering them in order to interfere with the making of clinical decisions. The HL7v2 and HLv3 versions of the HL7 standards are deemed obsolete as stated by the authors of the study [18].

The latest, and currently recommended, version of HL7 standards is FHIR. It is this version that will be evaluated with regards to SDAM later in the article. This new standard is on the rise. It is implemented not only in North America, but also in Africa (as described by the authors in [19]), and in Europe (to connect to a i2b2 database, which is an open-source clinical data analysis platform) [20,21]. The HL7 organisation recommends putting FHIR into production jointly with the TLS protocol to guarantee confidentiality and integrity in data exchanges. Yet, communications security can pose risks when relying solely on TLS. These risks can range in nature from data leaks, mainly due to insider threats, routers or poorly configured firewalls, and web services [22]. More specific threats to TLS can be paradoxically explained by the reduced effectiveness of firewall-borne security mechanisms when TLS is used across them. The encryption between a client and a server makes Next-

Generation Firewalls (NG-FW) and Intrusion Prevention Systems (IPS) difficult to use. Security systems such as URL filtering, malware detection, IPS signatures, and all of the advanced network features based upon Deep Packet Inspection (DPI) are all affected by this decline in efficiency [23,24]. In response to this, managers tend to enable the intermediate TLS decryption option. When this TLS option (TLS1.2 or TLS1.3) is enabled, upstream or at the security device level [25,26], it becomes possible to recover total visibility over traffic flows and protect organisations from threats. The problem is that, by turning on TLS decryption, it allows for disclosure of where a connection originates [27] and for man-in-the-middle attacks [28], which are made even easier when exchanging highly sensitive data due to the fact that the network packets are in clear text within network equipment. Since intermediary entities can read data in plain text, the confidentiality and integrity requirements of exchanges is not met, which is potentially even more problematic.

It is possible to use proxies to hide data sources away from the eyes of malevolent viewers, as presented in this article [29]. Nevertheless, this technique calls upon TLS only, leaving the proxy server reading data in plain text, and creating an obvious risk for data confidentiality, as described above. Therefore, despite the use of a proxy and TLS, confidentiality and integrity requirements are not fully achieved.

In [30], the authors offer a protocol for sending data with end-to-end security. This technique is based upon a mix-network, in other words, a network where exchanged messages are mixed together at every relay node. These mix-networks use a protocol that protects metadata in messages: The Sphinx protocol [31]. However, each intermediary can decrypt the payload, which affects the overall confidentiality of the system.

Finally, blockchain developments over the last few years have inspired many health researchers to try to improve health data sharing using techniques that have the potential to bring integrity and audit properties into a single system. Some of the new approaches [32] are inspired by the multi-layer cryptosystems (e.g., Bitcoin [33]), also known as onion routing, to exchange information [34].

Other blockchains, like Ethereum, reference [35] hide the transactions that are recorded in each block by adding a workable computational load. Nevertheless, it is still feasible to verify and see the transactions of a block by recalculating it. Therefore, this obfuscation does not achieve the confidentiality requirement. Moreover, these Ethereum-based blockchains present the same underlying risks; namely, that the network topology can be known by performing a traffic analysis and therefore network packet metadata can be intercepted and read [36].

In order to avoid some network nodes from reading packets containing sensitive data, some blockchain frameworks, like the HyperLedger Fabric, create private channels to better support confidentiality [37]. Some healthcare projects [38,39] use this approach, combined with a public-key infrastructure, to protect access to stored data. This approach relies on the ability to protect the private key of the participants to maintain confidentiality. Notwithstanding the interesting security features proposed, there are recurring security issues like threats by traffic analysis on these systems [40]. The confidentiality of exchanges and data sources is continually threatened due to new possible attacks on the architecture or the cryptographic keys exchange protocol [41]. Finally, the ability to support data queries required by a LHS to minimise data exchanged and the capacity to identify a responsible organisation for the extracted data remain as challenges.

As illustrated above, no reviewed approach fulfils all communication and security requirements to fully support a LHS. The next section will present a framework to express activities and agents of a data access system as part of a LHS. This will enable the formalisation of the specific requirements and structure of SDAM.

3. Materials and Methods

A Data Access System (DAS) enables data to be accessed and processed by different entities, in different contexts, while enforcing applicable rules (legal, ethical, or technical). This section presents a framework to describe a DAS through its processes and entities

within the context of a LHS. A DAS framework, when coupled with the right modelling, for both activity workflows and communications, can enable a significant degree of automation and offer a desirable security profile. These benefits are especially important when applied to the support of a learning cycle as explained above.

3.1. Data Access System Activities

A DAS, when used within a LHS, can be divided in two high level activity groups; a data source integration group and a specific project activities group. Figure 1 shows the phases for a DAS and related activities. In this paper, we will focus data extraction and transmission, which are arguably the most complex and risky activities. A data source integration will occur once for a given source (with eventual updates if the source changes). Once integrated, it can be reused multiple times across various projects without requiring more work. The specific project activities can be represented through two phases: Project approval and the project execution itself.

The integration of a data source includes both the mapping local data structure to a central model, to enable distributed queries, and individual indexing. Individual indexing involves sending demographics to a master indexer and receiving a source code for each individual represented in the data source. This code will later be used to coordinate the generation of project specific identifiers. While multiple methods to achieve individual indexing have been proposed, a detailed description of them is beyond the scope of this article.

The integration activities need to occur only once and when completed, the data source is available to participate to projects. As a last step, the data source can be registered to a repository to make its availability publicly known, if desired.

Assuming that required data sources are integrated and available, specific projects can be designed. The project artefact provides all the information required to execute a project. It includes the Project Workflow (PW), but also other parameters such as the list of participating organisations (and means to confirm their identities).

The PW provides information about the activities, their sequencing, and the required parameters to orchestrate progression through the workflow (input and output parameters, conditions for transitions, etc.). The PW also references data related information, such as queries to be executed on sources or data transformations to be applied.

In this framework we categorise information exchanged into two groups. The first one is referred to by result sets. They are generated by executing queries on data sources. The second group comprises Project-Specific Identifiers (PSI). These synthetic identifiers are generated at the start of each project so that the entities can reference individuals.

After coordination with the different organisations envisioned to be participants in a project, the project artefact (including the PW) is circulated for approval. Assuming the project is acceptable to all participants, they each need to sign the artefact thereby, confirming that they are intent on executing the activities assigned to them. The signature is performed using an encryption key. This is done to ensure the integrity and non-repudiation of the parties. Given the pragmatic problem of the ordering of signatures, the project initiator is responsible for making sure that all the signatures are present after which the artefact can be submitted for broadcast. At this point, all involved receive a fully signed copy making the project official and eligible for execution.

Here is a list of entities modelled for project orchestration with brief definitions:

- Anonymiser: Entity responsible for translating a set of PSI known to one entity into a set known to a different entity;
- Execution Engine (EE): Entity responsible for processing the PW and coordinating activities across all participating entities;
- Data Connexion Entity (DCE): Entity with an access to a data source (e.g., database, flat files, etc.). The DCE communicates with the EE. The DCE is responsible for executing the data activities using its connected data source;

- Evaluator: Entity responsible for receiving PSI sets from DCEs in order to create a cohort of individuals fulfilling some criteria. The information required to decide if an individual is to be included or not might come from different sources. Using PSI sets from all involved data sources, the evaluator executes operations to combine the source PSI sets and generate a final PSI set representing the individuals fulfilling the cohort criteria;
- Result Connexion Entity (RCE): Entity responsible for receiving result set blocks and PSIs from DCEs. The RCE then generates a unified final dataset linking data elements pertaining to a given individual.

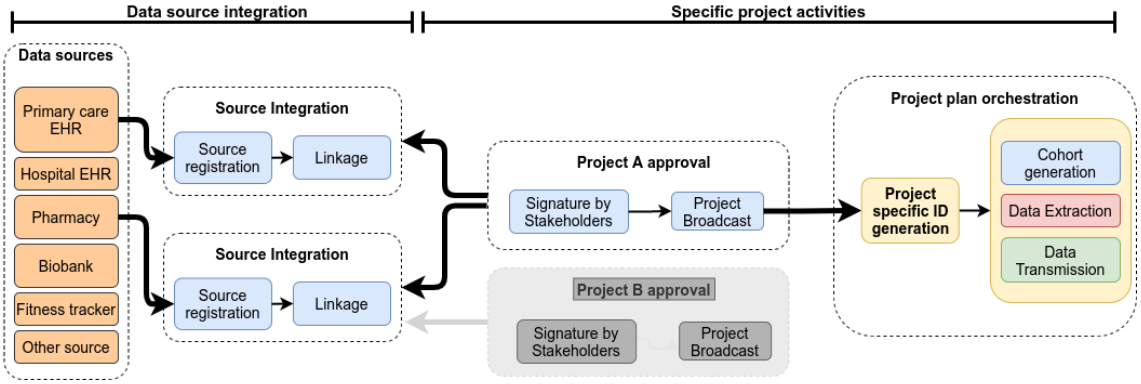

Figure 1. The different activities for a Data Access System (DAS) and their interactions. The diagram shows data sources with the three phases, namely source integration into the ecosystem, project approval, and activity orchestration. The flow of activities within the orchestration phase may vary depending on the project workflow.

Once the project artefact is broadcasted, the plan orchestration phase can begin. The first step is to generate the PSI sets. Using the individual source codes, the master indexer and the anonymisers will collaborate to generate and distribute PSI sets to each entity requiring them, each getting a different set. To be able to communicate about an individual, two entities will therefore need to communicate through the anonymisers in order to "translate" their PSI into one known by the indented receiver. If, for example, a project EE needs to instruct a DCE to extract data about a specific individual, the EE will send its PSI for that individual to an anonymiser who will translate it in terms of the DCE PSI and send it to the DCE.

Once all identifiers are in place, the PW execution itself can begin. For projects involving health data usage, three high-level activities type are commonly and can be described as follows:

1. Cohort generation. Using a set of criteria used to create a cohort of individuals. This results in a list of PSI that represent all individuals fulfilling this set of criteria. It can subsequently be used as a parameter to guide PW activities. Since criteria might need to be evaluated across multiple sources (for example, a cohort requiring a certain genomic trait based on data in a biobank and exposure to a certain medication based on electronic medical record data), the evaluator is responsible for combining logically the received sets of PSI to generate the final list of PSI representing that cohort. Since the evaluator does not share its PSI set with any of the DCE, communications with DCEs need to go through an anonymiser for translations;
2. Data extraction. When a DCE receives the command to proceed to a data extraction for a given cohort, it: (a) Connects to its data source, (b) uses the query indicated in the plan to extract data, and (c) uses the cohort provided as a parameter to restrict the result set to the individuals part of the cohort. The result set is now local to the DCE and can be used for further processing (like pre-loading a web form locally) or can be transmitted;

3. Data transmission. Data transmission is not automatically triggered by a data extraction. The result set could be used exclusively by the DCE internally. In addition, the transmission could be conditional to some other action as formalised in the PW. The transmission of a result set by a DCE to a RCE implies two communication streams. First, the result set needs to be sent. Second, the PSI list also needs to be sent to enable the RCE to relate various data blocks pertaining to the same individual.

It should be noted that any of these activities can be referenced multiple times in a PW.

As part of data transmission, a result set is sent from the DCE to the RCE, but since the RCE and the DCE do not share the same PSI set, the PSI need to take a different path. The PSI will be sent from the DCE to anonymisers for translation into the final data set to be received by the RCE. This translation enables the latter to piece data together.

At this point we can identify three broad categories of messages, or communications: (a) Commands/statuses (e.g., from the EE to DCE to instruct it to perform a data transmission, or from a DCE to the EE confirming that the transmission has been successfully completed), (b) result set transmissions (from DCE to RCE), and (c) PSI set transmissions.

3.2. Detailed Communications Sequences for Data Related Activities in a DAS

In a DAS, different activities will require different messaging sequences. The execution engine is the entity controlling the flow of activities and as a result, the sequences illustrated in this section start with an instruction from the EE. Figure 2 offers a step-by-step depiction of the extraction and transmission activities (including all three types of messages: Commands, extracted result sets, and PSI). The sequences are designed to minimise data exchanges, including communications failure (e.g., do not send a list of PSI for a cohort before you have confirmation that the DCE is available and ready to execute the data extraction). Some simplifications are used here. For example, many health organisations will not allow direct contacts as mentioned previously. A DCE would therefore "ping" the EE through its proxy to request instructions, rather than the EE directly communicate with the DCE. In addition, sub steps like handshakes or activity completion acknowledgements are not represented.

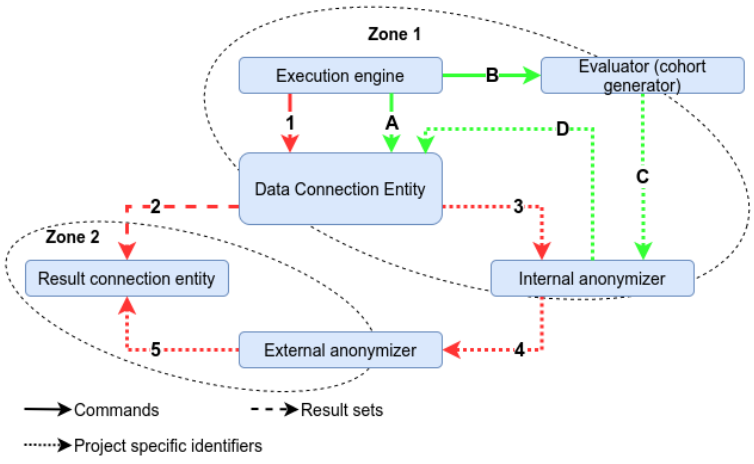

Figure 2. Model of Data Access System. (Green) messages used during an extraction activity. (Red) messages used during a transmission activity.

Of note, there are two anonymisers meant to segment PSI data. In order to ensure a maximum amount of redacted information available, while curbing the risk of re-identification, the framework is segmented into zones. Zone 1 is the internal area. It groups the DCE together with the project entities not handling result sets (e.g., execution engine, etc). Zone 2 is the so-called external area, including entities allowed to receive result sets, i.e. , the RCE. The risks and allowed operations differ in each zone, hence the use of in-

ternal and external anonymisers to channel PSI in different zones. The internal anonymiser is required for communications involving an entity in zone 1 and the external anonymiser is required for communications involving entities in zone 2. Therefore, a communication between zone 1 and 2 (e.g., data transmission) will require both anonymisers.

An extraction activity requires the following messaging steps: (A) The EE sends a command with a reference to an extraction activity in the plan to a DCE. (B) The EE sends a command to the evaluator instructing it to send the list of PSI for a given cohort to the internal anonymiser. (C) The evaluator sends the PSI set for the cohort to the internal anonymiser. (D) The internal anonymiser sends the translated PSI set (using the DCE set) to the targeted DCE.

A result set can be transmitted to an RCE for downstream usage by the following steps: (1) The EE sends a command with a reference to a transmission activity in the plan to a DCE. (2) The DCE will send messages to the RCE, mentioned in the plan, to transmit the result set (as a set of data blocks). (3) The DCE will send the PSI list associated with this result set to the internal anonymiser. (4) The internal anonymiser sends the translated PSI list (using the external anonymiser PSI set) to the external anonymiser. (5) The external anonymiser sends the translated PSI list (using the RCE PSI set) to the targeted RCE. This communication flow ensures that the result set itself and the associated PSI list are not transmitted through the same path. The method of keeping the data blocks linked with the PSI in this transmission lie within a specific protocol in the suite, which is outside the scope of this article. Suffice to say the information required to piece everything together at the RCE is part of the messages exchanged.

3.3. Functional Security Requirements

As mentioned previously, existing approaches do not comprehensively fulfil all the high-level requirements mandated by LHSs. More specific security requirements, to be handled by a LHS communication protocol suite, can now be formalised in SDAM using the DAS framework:

- Multilevel encryption: Encryption is essential for the confidentiality, authenticity, and integrity of messages exchanged. Each layer must use unique dedicated keys. Different cryptosystems can be used in order to adapt to each layer when needed. When encryption is on, message contents become readable to source and destination entities at the presentation and application layers. The Encrypt-then-MAC method is used for integrity purposes;
- Data confinement: The application layer must be the only layer that can insert and handle content between two entities. Intermediate entities are not able to access the message payload in clear;
- Registration and authentication: Organisations participating in the project and are responsible for entities like DCE or RCE, need to be formally identifiable in the project artefact (for example, by a certificate authority). The project artefact needs to be signed by all involved. Communication protocols should only make use of validated information;
- Resilient anonymous routing: SDAM must use its own routing system between entities, thus reinforcing the datalink layer routing map. The bridge between the routing system of SDAM and the IETF protocols must be established only at the network layer. In addition, SDAM cannot require public disclosure of the source and destination entities IP addresses. As a result, an entity shall never connect to another one in a direct manner. They are instead required to go through a pool of proxies (of size three to n). Consequently, there is nothing deterministic about the paths between the two entities in cases where the pool is strictly larger than 6. For this to work a proxy must offer a public IP address;
- Self-adaptive protocols: Different types of communications will require different performance profiles. For example, a result set transmission between a DCE and an RCE requires a higher throughput, while sending commands demands lower

latency. SDAM must therefore be flexible enough to fine-tune the speed and latency for different message types, volume, or frequencies;

- Logging and traceability: All the actions performed at each layer are logged. This ensures non-repudiation for external audits. The messages exchanged and the connections between the entities are logged. These log entries should be sent to an external audit log to maximise security benefits;
- Signature and certificate authorities: Entities, messages, and packets, as well as log entries are signed and certified. This achieves the requirement of authentication. It also provides two additional benefits pertinent to security in a broader sense: Auditing and non-repudiation;
- Validation procedure: A communication protocol suite must offer a system to validate and verify layer functioning. For example, requested activities must go through checking so as to ensure they are part of the project workflow;
- Standardised layer interfaces: Layer interface specifications are standardised. Standardisation leads to improved interoperability and, in the event of changes in exchange model, or the introduction of new protocols, maintainability of the system. In addition, this standardisation provides additional design security [42].

3.4. Sensitive Data Access Model for Healthcare

This section presents the Sensitive Data Access Model (SDAM) for LHSs, starting with the assumptions used to develop it and the key points of the new architecture. Followed by a description of the model layer, highlighting important features. Ending with a discussion of the principle of anonymisation of the IP addresses using proxies between the entities during the extraction and transmission activities.

3.4.1. Preliminaries

The assumptions upon which SDAM is based can be summarised as follows:

- Assumption 1 is about the overall infrastructure, assuming that communication lines are not secure and can be tapped;
- Assumption 2 holds that a message can be intercepted and modified by anyone at any time as early as it is sent out by a source entity up until it reaches the destination entity;
- Assumption 3 takes into account that entities can be compromised. A SDAM compatible protocol suite should be designed in such a way that a maximum number of entities (other than DCE and RCE) would need to get compromised before a significant risk of associating pieces of data to an individual outside permissible contexts would occur. This underpins the principle of defense-in-depth;
- Assumption 4 holds that entities will perform cryptographic operations such as hash functions and some other forms of encryption when required.

3.4.2. Key Architectural Principles

Figure 3 shows data flow between two entities and associated proxies through various layers. SDAM allows for exchanges between entities on the application, presentations, sessions, and transport layers. As for the network layer, these are node-to-node exchanges. The source entity connects to an entry proxy; from that, the message goes on through a pool of proxies (thus helping to keep the actual path hidden) and then the message ends up reaching an exit proxy linked to the target destination entity. Proxies are blind to message contents given the encryption applied. Since the routes and types of message (encrypted) transiting by the proxies change regularly, even a compromised proxy poses a minimal threat in terms of traffic analysis to enable message payload recovery.

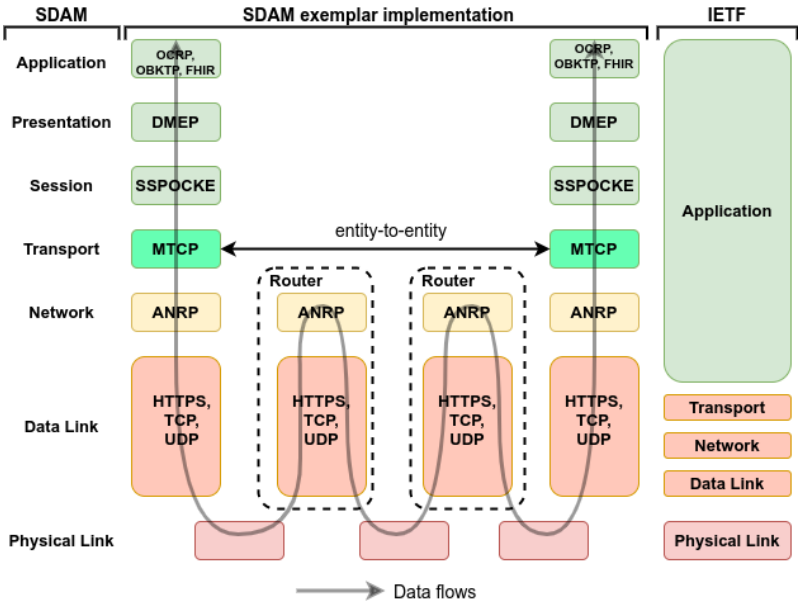

Figure 3. Sensitive Data Access Model (SDAM) exemplar data flow. SDAM layers are represented on the left and their corresponding representation in the IETF stack can be found on the right. The middle part illustrates an exemplar data flow with candidate protocols for the different layers.

3.4.3. SDAM Layer Descriptions

SDAM has specific features for each of the different layers included in the model. These features are described in the following sections. The PARS3 [43] data access system, in deployment in Canada [44], is currently using SDAM to structure its communications. Candidate protocols from the project, and other standards like FHIR, are mentioned in the different layers (a full presentation and discussion of these protocols is out of scope for this article).

IETF Layers Implemented in the SDAM Data Link Layer

The IETF reference model is used in SDAM in order to structure the low-level communications. As mentioned in the high-level requirements, this reference model is used because it is widely implemented, which makes interoperability easier across multiple organisations in different context. The SDAM data link layer is concretely implementing the IETF Data Link, Network, Transport, and part of the application (for HTTPS) layers. It uses existing protocols, which can be high-level like HTTP or HTTPS or low-level like TCP or UDP.

While SDAM's use of IETF is very beneficial in terms of interoperability, the IETF reference model by itself does not meet the security requirements outlined in this article. SDAM answers this challenge by structuring additional OSI based layers on top of the SDAM data link layer, enabling flexible implementations of defense in depth principles. These SDAM layers will now be presented starting with the application layer.

Application Layer

The application layer is in charge of supporting the application services, which manage the messages and their processing. This layer uses high-level, entity-based routing to define the recipient of messages. As per a project workflow, the application layer must ensure that the entities comply with what has been defined and validated by all the project stakeholders. It is through the application layer, for example, that the PSI translation service of an anonymiser would receive and send the PSI sets.

The application layer is heavily involved during data extractions and transmissions. These exchanges of PSI and result sets require a protocol that will maintain the integrity and

the full semantic of the data transmitted. HL7 FHIR could be leveraged in this layer as could the Ontology-Based Knowledge Transfer Protocol (OBKTP) used in the PARS3 system.

In order for the PW to be executed as planned, entities must be able to exchange messages containing commands (and statuses confirming availability and executions). The Operation Command Relay Protocol could be used for these exchanges. Communications from the application are passed to, or received from, the presentation layer.

Presentation Layer

The presentation layer offers several features related to the restructuring of data received in the application layer, and vice versa. The presentation layer deals with encoding data in a particular format and encrypting the content, similarly with decryption. A protocol on this layer can be used to provide a symmetric encryption service for message content that is enforced using a Symmetric Data Encryption Key (SDEK). It is important to note that this protocol can also take care of slicing and splitting both result sets and PSI sets into small chunks, a requirement of the DAS framework. Then, it can encrypt all of the content of a given type that will be sent to the destination entity. Only the recipient's presentation (responsible for the decryption) and application layers are able to read this content. A protocol here must manage encryption keys, which can be entrusted to local hardware, such as, Trusted Platform Modules (TPMs), or remote hardware or software security modules, such as Hardware Security Modules (HSMs). Candidate models include the Data Mixing and Encryption Protocol (DMEP).

Session Layer

The session layer is meant to synchronise communications between two entities. It is also used to restore previous exchanges in the event of errors. A protocol on this layer is in charge of creating a session and keeping connections open among entities. In addition, it must guarantee non-repudiation between the identities of the sender and the recipient, as well as the integrity of the data when sent. The presentation layer sends and receives the messages from the session layer. A protocol of this layer must offer certain services, such as the recovery of the public keys of the entities with which they communicate and perform asymmetric encryption operations. When a session is created, the source entity will generate a session token and, during the initial exchange, Symmetric Key Agreement Keys (SKAK) are exchanged between the entities to share the SDEK key. These different exchanges make it possible to have encrypted and authenticated messages within the framework of a session. The Secured Session Over Cryptographic Key Exchange (SSPOCKE) protocol was conceived as part of the PARS3 development to cover this layer.

Transport Layer

Data flow control and reliability of communications between entities are managed by the transport layer to ensure that the destination entity receives exactly what was sent from the source entity, in the same order. Optimising transport with numbering and reorganising segments, as well as adding extra mitigation methods against some attacks are done here as well. The transport layer's objective is to ensure that the data to be sent to the network layer will be of an appropriate size, or else fragmented into several packets. The transportation layer also provides a means of implementing systems of detection against advanced DDoS attacks. It is with this in mind that the Message Traffic Control Protocol (MTCP) has been designed in PARS3.

Network Layer

The network layer deals with routing, relaying, and logging. This layer bridges the IETF standards implemented in the SDAM data link layer (presented above). Routing means finding a path among multiple entities and meeting the project workflow requirements. This is achieved by relaying messages via intermediate nodes (proxies). The network layer manages the proxy pool and the dynamic topology, which is core to SDAM.

The network layer features the integration of proxies to protect the anonymity of entities, the use of a dynamic logical topology, end-to-end encryption, and a method of batch processing of the messages. These security requirements and mitigation methods will help protect this layer from a wide variety of traffic analysis threats. Network layer datagrams are made of metadata and content (payload). The metadata are session ID, log code, and path.

The network layer also logs incoming and outgoing messages. Each intermediate node sends exit and entry notifications to the log and then delivers the message or response to the next node. The target destination entity will transmit the content to the upper layer and then transmit a notification in response to the source entity. The PARS3 Anonymous Network Routing Protocol (ANRP) is a candidate for this role.

3.4.4. Application to Data Activities

One of the basic pillars of SDAM is to guarantee that entities remain anonymous, in terms of network addresses, to each other and that there is no direct communication between them. For this purpose, a pool of proxies and anonymisers have been integrated into the architecture as part of the network layer.

Proxies are meant to transfer messages between two entities while not requiring a public record of the source or target network IP address. In addition, the proxies constitute a line of defence. They help dodge traffic analysis attacks on multiple sources. These benefits require that an entity chooses one or many trusted relay proxy when participating in a project. All the proxies of a proxy pool attributed to a project are registered in the project artefact and so, are verified and can be validated by any participant. It is important to note that, a proxy cannot receive both PSI and result sets simultaneously. This is to avoid the concurrent existence of result set and PSI messages on the same proxy (which could increase the re-identification risk). The SDAM proxy pool is characterised as an enhanced mix-network with dynamic topology. Its applications to the extraction and transmission activities are detailed below. A discussion of the parameters and functions used to dynamically change the proxy topology would be part of a protocol implemented for the network layer, and, as such, are beyond the scope of this paper.

A snapshot of extraction exchanges is described in Figure 4, given the dynamic nature of the topology. The emphasis here is on the added degree of details in message paths (compared to the high-level view presented previously) by including the proxies. Data extraction communications require command and PSI messages. During the period illustrated by the figure, a given proxy will only handle one type of message or the other. It should be noted that, the number of proxy-to-proxy exchanges and the exact path taken will change for each communication. The internal anonymiser must be called upon to translate the PSI set representing the cohort to be used, since it involves a transmission exclusively in Zone 1.

Once the extraction activity comes to an end, and it is instructed by the EE to do so, the DCE can launch a transmission activity. Figure 5 shows the required messages to transfer the result set from the DCE to the RCE with the required PSI. These exchanges make use of the internal and external anonymisers (given the need to cross both zones) and the proxy pool. The execution engine will also trigger the activity with a command message, but it is omitted here for readability. The proxy pool principles are the same, with the emphasis on PSI and result set types of messages. In the end, the RCE received, through different paths, both the result set data and the related PSI enabling it to regenerate the unified result set ready to be used.

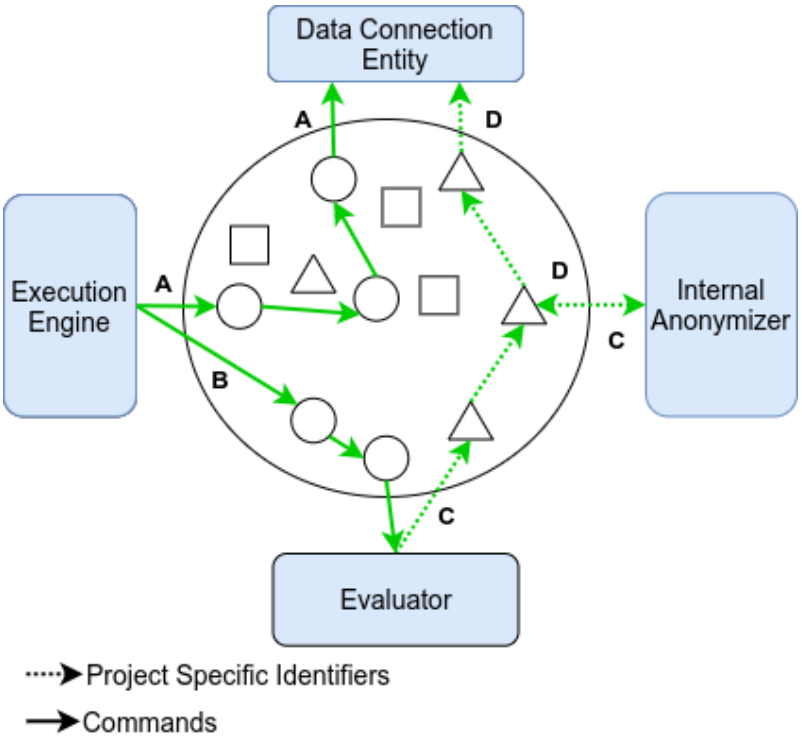

Figure 4. Exchanges in an extraction activity with proxy usage. Circle proxies handle communications data whereas triangle proxies process attribute data at a given moment in time.

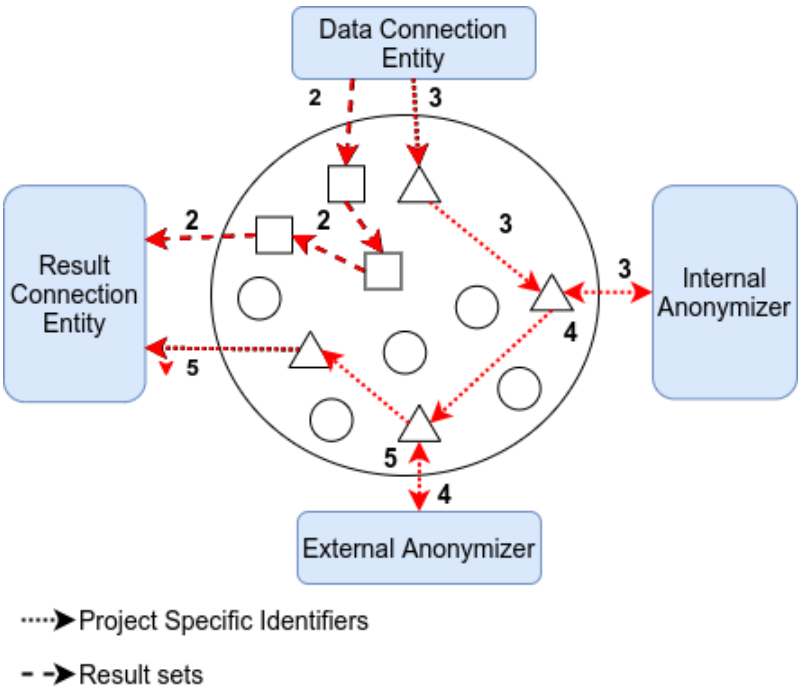

Figure 5. Exchanges in a transmission activity with proxy usage. Both result sets and project specific identifiers are being isolated and segmented. Square proxies handle result sets whereas triangle proxies process project specific identifiers.

4. Results and Discussion

This section will present an exemplar application, leveraging SDAM: Reflex-D accessing data through PARS3 which follows SDAM. It is used to provide clinicians with insights into how their patients with diabetes are cared for (types of treatments, aggressiveness for

reaching targets, follow-up frequencies, interactions with other drugs, etc.). What follows is a security properties and vulnerability analysis of the SDAM design. This analysis is presented in comparison with FHIR (with and without TLS).

4.1. Reflective Pratice for Patients with Diabetes

The SDAM was followed to enable data access by an audit and feedback tool made to improve care of patients with diabetes by providing insights to their primary care physician. The tool, named Reflex D, is an application for auditing and analysis. It is a tool for reflection and analysis available to healthcare teams to engage in a patient-centred approach [45]. These teams can use the tool to learn from the trajectories of their patients, highlighting important facts and evaluating their capacity to follow best practices.

Figure 6 shows the evolution over time of laboratory test results along with medication history for HbA1c and creatinine.

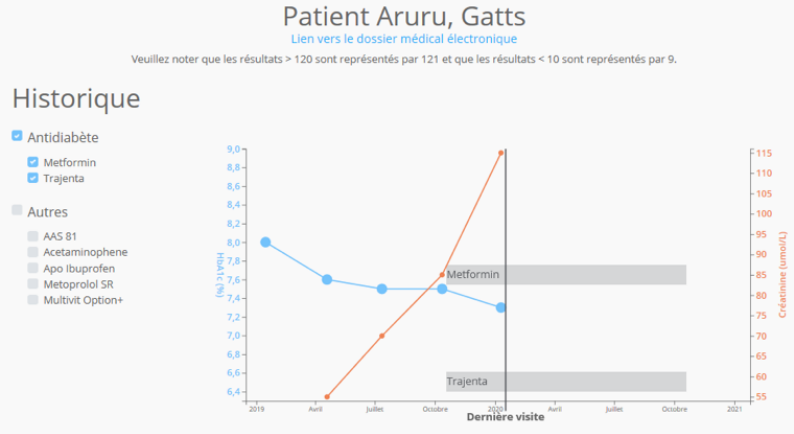

Figure 6. A sample profile of a patient showing medical history and records of HbA1c and creatinine developments.

Similarly, relevant practice guidelines from national institutes or professional organisations are contextualised to enable optimal care. Figure 7 shows that the latest test results can be used to determine if any of a patient's active drugs are contraindicated and, subsequently treatments can be adjusted.

To achieve the results mentioned, the tool named ReflexD needs to be able to extract the right data for diabetic patients for each clinician. While the tool could use data coming from any setting, where the required information is available, Reflex-D has been developed specifically with primary care clinicians in mind. Having access to a primary care Electronic Medical Record (EMR) is challenging in many ways, as these tend to be smaller organisations with lower information technology resources [46]. Trust needs to be built. Security is of the upmost importance, since the EMR often contains longitudinal data over decades. This includes, not only data about care happening at the clinic, but, often times, summaries of care episodes that have occurred in hospitals for example. Being able to use SDAM to guide the implementation of a sound communication protocol suite is therefore extremely useful in such a setting. Once connected, the same communication approach can also be used to participate in clinical research project as well. The use across various medical platforms, provides a higher return on the initial investment of source integration in a broad LHS context, instead of in silos for each tool (see Figure 1).

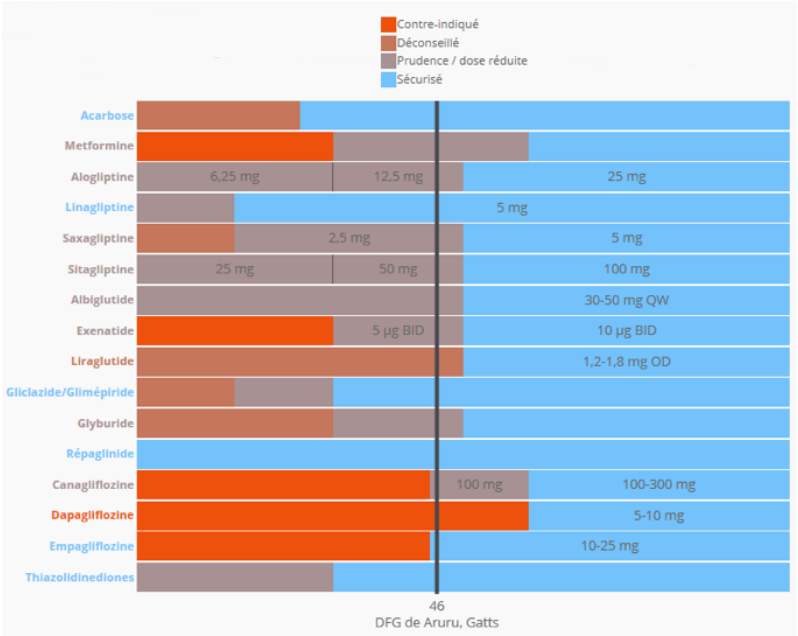

Figure 7. Automated analysis of current renal function for a patient and visual indications to help choose a safe diabetes drug.

4.2. Security Properties Description

In view of the LHS's security requirements, presented previously, a description of security properties provided by adhering to SDAM follows. Moreover, given the preponderant role FHIR currently plays in structuring health data exchanges, we will also evaluate if similar properties are present when using FHIR with or without the concurrent use of the TLS protocol.

Confidentiality is a cornerstone when handling sensitive data. SDAM provides distinct authentication mechanism properties. The protocol in its presentation layer is responsible for encrypting data from entity to entity. It does not enable intermediate nodes to have access to the payload in clear. The use of FHIR over TLS adds encryption between communication relays, but still allows for the possibility of a node decrypting the payload. The result is that the confidentiality criteria are not fulfilled.

In SDAM, the integrity of messages is ensured by authenticated signatures. FHIR users can be certified using a single-level authentication server [47]. Refresh tokens could be compromised when in transit. Conversely, SDAMs checks and validates users' IDs at the plan (which has been signed) and node levels.

In order to provide trust to participants, traceability and logging are essential. Traceability is meant to identify the origin and route of commands, as well as the different types of data that pass between entities. These traces must be recordable by an external audit entity and kept accessible in a location controlled by a trusted party. Logging refers to the means by which to register this information. SDAM requires that protocols should have the possibility of logging on all entities, even those close to data sources such as the DCE or outside the system such as applications connected to the RCEs. In addition, each action, on each layer, should be recorded by the network layer on the local system and simultaneously be sent to an external log manager. It should be noted that these logs can be encrypted too.

FHIR uses two resources, known as "Provenance" and "AuditEvent", in order to keep up with the filling of the log. Nevertheless, FHIR is not designed to inherently support an outsourced log management provider. The connections to external logs would need to be specified, developed, and implemented by a party wanting to achieve this, which affects interoperability and potentially, the overall event tracking profile.

The purpose of calling on external logs is to have an audit tool with input actions. However, it is also meant to provide another independent communication channel so that entities can factually report any malicious activity in the event of attacks. SDAM relies on validated third-party certificates, a robust encryption key infrastructure, and signed project artefacts to ensure non-repudiation. To input new events, FHIR relies on a resource called "Provenance.signature", which is a digital signature. Using a single digital signature is problematic because it does not achieve non-repudiation. This is because there is no evidence that the signature does belong to an entity.

SDAM includes cryptographic handshakes, specifically handled by SDAM. It also leverages IETF standards, when possible, like HTTPS. In both cases, a third-party authority is responsible for distributing proper certificates and ensuring that the organisations are fully authenticated. FHIR does not provide point-to-point authentication. Communication security must be ensured by the TLS protocol.

Table 1 summarises the analysed security properties as found in the LHS security requirements and FHIR.

Table 1. SDAM security properties and evaluated in context of Fast Healthcare Interoperability Resources (FHIR) with and without TLS use.

Security Property	FHIR w/o TLS	FHIR w/TLS
Confidentiality	×	×
Integrity	×	○
Authentication	○	○
Traceability	○	○
Logging	○	○
Non-repudiation	○	○
Authenticated point-to-point link	×	✓

✓ Supported ○ Partially supported × Not supported.

4.3. Security Vulnerability Analysis

The second analysis consists of an evaluation of the suggested means of defences of SDAM against the most common threats on communications and data access protocols, as outlined in the LHS security requirements. A comparison will be provided for the same threats when using FHIR.

SDAM offers mitigation responses to masquerade attacks by adding verification and integrity seals sequentially as the transmission progresses. Attacks of this kind can be detected by exploring logs and validating identities. FHIR could provide some mitigation by imprinting the messages with an embedded certificate. The use of FHIR with TLS does offer a fair degree of protection at that level.

Replay attacks could have significant consequences in a context of decision support, for example. One of the ways SDAM mitigates this type of attack is by providing a session number and encrypting the message content at the presentation layer, keeping the message undisclosed to a third party while in transit. This aspect has not been addressed by the security mechanisms proposed by FHIR.

Traffic analysis, which is an instance of side channel attack, is a difficult security issue to tackle. This is why SDAM goes to great lengths in limiting its attack surface as much as possible. SDAM prescribes the use of a proxy pool using a dynamic topology. This avoids leaking information about the volume and communication frequency. FHIR does not offer a defence in its design to defend itself against these attacks.

Table 2 summarises the vulnerabilities addressed by the SDAM and the protection mechanisms offered by FHIR.

Table 2. Vulnerabilities addressed by SDAM and an analysis in context of FHIR with and without TLS use.

Vulnerability	FHIR w/o TLS	FHIR w/TLS
Masquerade attacks	×	✓
Message Replay	×	×
Trafic analysis	×	×

✓ Secure ○ Partially secure × Not secure.

As can be seen, SDAM delivers a favourable security profile, based on its design, and addresses the most critical types of attacks in context of a LHS. The Denial of Service (DoS) attack has not been evaluated, as it is linked significantly to the lower parts of the communication infrastructure and will depend on specific implementations. Nevertheless, we can mention that not requiring the publication of entities IP addresses and the possibility of using multiple trusted proxies within a large pool will provide some protection.

It is important to note that FHIR has been included given its use in multiple contemporary projects. Nevertheless, it does not advertise, nor claim, an extensive security profile covering multiple layers (coherent with the name of its parent organisation, HL7). It is still important to convey the message to the LHS community that using HL7, even with TLS, does not address all the requirements identified.

5. Conclusions

Learning health systems have been widely regarded as the way forward to bridge the gap between scientific discoveries and impacts on individuals' health. It should be noted that, while LHS have been developing for more than 15 years, an evaluation to provide a comprehensive security requirement model was missing. This article provides such a model, SDAM. High-level requirements, including addressing confidentiality, deidentification, data linkages, and mitigation against common threats, as well as leveraging, as much as possible, widely-used standards, like the ones from IETF, are critical to the support of secure, acceptable, and regulation-compliant LHSs at scale. Currently implemented approaches, like those proposed by HL7 (including FHIR), as well as newer, more experimental ones, like projects based on blockchain technologies, do not cover all necessary requirements but can certainly contribute to the solution. For example, the approaches using blockchain technology are primarily turned toward audit purposes. To this end, it would be interesting to use blockchains as a general ledger to store the information logged from the different components and protocols used by LHS. This could ensure the integrity and non-repudiation of the various activities on a platform between several stakeholders from multiple organisations.

When designing SDAM, specific security requirements were taken into account. OSI was used as a basis for the layers and IETF standards were leveraged as much as possible. Traffic analysis is a significant threat and is difficult to protect from. Given the requirements to maintain network address anonymity for the entities, particularly the data sources, great care has been put in designing the proxy pool requirement as an enhanced mix-network with a dynamic topology. To illustrate the applicability to concrete LHS activities, an exemplar use-case was illustrated and SDAM advantages discussed. We ended our discussion with a comparison of the security properties offered by SDAM compliant implementations and, an evaluation of their presence or absence within FHIR. A similar exercise was proposed with regards to security threats and mitigation strategies offered. While it would be false to state that FHIR advertises, or claims, that it covers the broad spectrum of requirements covered by a model based on the full OSI stack, it is important to make explicit to the LHS community that relying solely on FHIR (with or without TLS as suggested by HL7) would pose significant security risks. While SDAM can prove useful for evaluating the suitability of communication protocol suites to support LHSs, work

remains to be done in order to produce and publish a suite of protocols that would be fully compliant with the proposed model.

From the definition of a learning cycle to its execution, including the transit of data through infrastructure layers, this work has opened avenues regarding the feasibility of improving data access between multiple organisations, while ensuring data security. Previous work demonstrated that significant gaps remain when security evaluations are envisioned on a per component basis rather than assessing security properties holistically at the system level. The new SDAM model inherently unifies data exchange formalisation during a learning cycle with the need for formal verifications of security criteria, which must be guaranteed on all stages of data-related activities. While SDAM provides a clear roadmap for desirable properties, ensuring its continued fulfilment remains a significant challenge: As systems change and topologies evolve, automated or semi-automated methods would help in ensuring traceability of security properties.

PARS3 is a data access system being developed to support multiple LHS deployments (mandated by the Quebec Health Ministry to support primary care data exchange in Quebec, the Ensemble network—a France-Québec rare disease network, etc.). As part of its development, protocols have been structured according to SDAM and have been mentioned earlier in the article. Their positive formal evaluation, in light of SDAM, could provide a first all-encompassing suite of protocols deployable in the field. However, trust is not dependant only on technological approaches. Transparency and consent management are essential in building an ecosystem of mutual trust among patients, clinicians, and the various stakeholders in LHSs. This implies that work will be needed to design protocol specifications allowing communications with new entities, like citizen portals, to allow them to express consent, preferences, or to provide them with transparency related information.

Author Contributions: Conceptualization, T.E., B.F., L.L., J.-F.E. and M.M.; formal analysis, T.E., B.F., L.L. and J.-F.E.; funding acquisition, J.-F.E.; investigation, T.E., L.L. and J.-F.E.; methodology, T.E., B.F., L.L. and J.-F.E.; project administration, J.-F.E. and L.L.; validation, B.F., L.L. and J.-F.E.; writing—original draft, T.E.; writing—review & editing, T.E. and J.-F.E. All authors have read and agreed to the submitted version of the manuscript.

Funding: This work was supported in part by the Unité de Soutien SRAP du Québec, Health Data Research Network Canada with the Canadian Data Platform and the Centre interdisciplinaire de recherche en informatique de la santé de l'Université de Sherbrooke (CIRIUS).

Acknowledgments: We are grateful to our colleagues for revising the analysis and the general architecture principles, and to the reviewers' insight. It greatly enhanced the scientific value of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SDAM	Sensitive Data Access Model
DAS	Data Access System
PW	Project Workflow
PE	Plan Entity
DCE	Data Connection Entity
RCE	Result Connection Entity
OBKTP	Ontologies-Based Knowledge Transfer Protocol
OCRP	Operation Command Relay Protocol
DMEP	Data Mixing and Encryption Protocol
SSPOCKE	Secured Session Over Cryptographic Key Exchange
MTCP	Message Traffic Control Protocol
ANRP	Anonymous Network Routing Protocol
SDEK	Symmetric data encryption key
SKAK	Symmetric key agreement key

References

1. Harrison, E.; Holdsworth, R. How many claudicants should be prescribed statins? *Eur. J. Vasc. Endovasc. Surg.* **2003**, *25*, 367–368. [CrossRef] [PubMed]
2. Sparrow, R.T.; Khan, A.M.; Ferreira-Legere, L.E.; Ko, D.T.; Jackevicius, C.A.; Goodman, S.G.; Anderson, T.J.; Stacey, D.; Tiszovszky, I.; Farkouh, M.E.; et al. Effectiveness of Interventions Aimed at Increasing Statin-Prescribing Rates in Primary Cardiovascular Disease Prevention: A Systematic Review of Randomized Clinical Trials. *JAMA Cardiol.* **2019**, *4*, 1160–1169. [CrossRef] [PubMed]
3. Institute of Medicine; Roundtable on Evidence-Based Medicine. *The Learning Healthcare System: Workshop Summary*; The National Academies Press: Washington, DC, USA, 2007. [CrossRef]
4. Cumyn, A.; Barton, A.; Dault, R.; Cloutier, A.M.; Jalbert, R.; Ethier, J.F. Informed consent within a learning health system: A scoping review. *Learn. Health Syst.* **2019**, *4*. [CrossRef]
5. Kaplan, B. How Should Health Data Be Used?: Privacy, Secondary Use, and Big Data Sales. *Camb. Q. Healthc. Ethics* **2016**, *25*, 312–329. [CrossRef]
6. Osadchuk, M.A.; Osadchuk, A.M.; Kireeva, N.V.; Trushin, M.V. Legal Regulation in Digital Medicine. *J. Adv. Res. Law Econ.* **2020**, *11*, 148–155. [CrossRef]
7. Morrison, M. Research using free text data in medical records could benefit from dynamic consent and other tools for responsible governance. *J. Med Ethics* **2020**, *46*, 380–381. [CrossRef] [PubMed]
8. Schneeberger, D.; Stöger, K.; Holzinger, A. The European Legal Framework for Medical AI. In *Machine Learning and Knowledge Extraction*; Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 209–226.
9. Delaney, B.C.; Curcin, V.; Andreasson, A.; Arvanitis, T.N.; Bastiaens, H.; Corrigan, D.; Ethier, J.F.; Kostopoulou, O.; Kuchinke, W.; McGilchrist, M.; et al. Translational Medicine and Patient Safety in Europe: TRANSFoRm—Architecture for the Learning Health System in Europe. *Biomed. Res. Int.* **2015**, *2015*, 961526. [CrossRef] [PubMed]
10. Zimmermann, H. OSI Reference Model-the ISO model of architecture for open systems interconnection. *IEEE Trans. Communication (USA)* **1980**, *COM-28*, 425–432. [CrossRef]
11. Jessadapattharakul, R.; Prom-on, S.; Tanprasert, C.; Achalakul, T. Data exchange protocol for healthcare service in Thailand. In Proceedings of the 2015 Fourth International Conference on Future Generation Communication Technology (FGCT), Luton, UK, 29–31 July 2015; pp. 1–6. [CrossRef]
12. Vito, D.; Casagrande, G.; Bianchi, C.; Costantino, M.L. An interoperable common storage system for shared dialysis clinical data. In Proceedings of the 2016 IEEE EMBS International Student Conference (ISC), Ottawa, ON, Canada, 29–31 May 2016; pp. 1–4. [CrossRef]
13. Nalin, M.; Baroni, I.; Faiella, G.; Romano, M.; Matrisciano, F.; Gelenbe, E.; Martinez, D.M.; Dumortier, J.; Natsiavas, P.; Votis, K.; et al. The European cross-border health data exchange roadmap: Case study in the Italian setting. *J. Biomed. Informatics* **2019**, *94*, 103183. [CrossRef] [PubMed]
14. Swanink, R. Persistent Effects of Man-in-the-Middle Attacks. Bachelor's Thesis, Radboud University, Nijmegen, The Netherlands, 10 January 2016; p. 43.
15. Townsley, W.; Valencia, A.; Rubens, A.; Pall, G.; Zorn, G.; Palter, B. Layer Two Tunneling Protocol (L2TP). Technical Report, RFC 2661, August 1999. Available online: https://tools.ietf.org/html/rfc2661 (accessed on 26 February 2021).
16. Singh, A.K.; Samaddar, S.G.; Misra, A.K. Enhancing VPN security through security policy management. In Proceedings of the 2012 1st International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 15–17 March 2012; pp. 137–142.
17. Dameff, C.; Bland, M.; Levchenko, K.; Tully, J. Pestilential Protocol: How Unsecure HL7 Messages Threaten Patient Lives. Available online: https://acsweb.ucsd.edu/~mbland/pestilential_protocol.pdf (accessed on 26 February 2021).
18. Bender, D.; Sartipi, K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal, 20–22 June 2013; pp. 326–331. [CrossRef]
19. Baskaya, M.; Yuksel, M.; Erturkmen, G.B.L.; Cunningham, M.; Cunningham, P. Health4Afrika - Implementing HL7 FHIR Based Interoperability. *Stud. Health Technol. Informatics* **2019**, *264*, 20–24. [CrossRef]
20. Boussadi, A.; Zapletal, E. A Fast Healthcare Interoperability Resources (FHIR) layer implemented over i2b2. *BMC Med Informatics Decis. Mak.* **2017**, *17*, 120. [CrossRef] [PubMed]
21. Pfiffner, P.B.; Pinyol, I.; Natter, M.D.; Mandl, K.D. C3-PRO: Connecting ResearchKit to the Health System Using i2b2 and FHIR. *PLoS ONE* **2016**, *11*, e0152722. [CrossRef] [PubMed]
22. Suga, Y. Status Survey of SSL/TLS Sites in 2018 After Pointing Out About "Search form" Issues. In Proceedings of the 2018 Sixth International Symposium on Computing and Networking Workshops (CANDARW), Takayama, Japan, 27–30 November 2018; pp. 483–485. [CrossRef]
23. De Carnavalet, X.d.C.; van Oorschot, P.C. A survey and analysis of TLS interception mechanisms and motivations. *arXiv* **2020**, arXiv:2010.16388.
24. Baek, J.; Kim, J.; Susilo, W. Inspecting TLS Anytime Anywhere: A New Approach to TLS Interception. In Proceedings of the 15th ACM Asia Conference on Computer and Communications Security, Taipei, Taiwan, 5–9 October 2020; pp. 116–126.

25. Radivilova, T.; Kirichenko, L.; Ageyev, D.; Tawalbeh, M.; Bulakh, V. Decrypting SSL/TLS traffic for hidden threats detection. In Proceedings of the 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, UKraine, 24–27 May 2018; pp. 143–146. [CrossRef]
26. Sherry, J.; Lan, C.; Popa, R.A.; Ratnasamy, S. Blindbox: Deep packet inspection over encrypted traffic. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, New York, NY, USA, 17–21 August 2015; pp. 213–226.
27. Frolov, S.; Wustrow, E. The use of TLS in Censorship Circumvention. In Proceedings of the NDSS, Network and Distributed Systems Security (NDSS) Symposium 2019, San Diego, CA, USA, 24–27 February 2019.
28. Automatization of MitM Attack for SSL/TLS Decryption. Available online: https://dspace.vutbr.cz/handle/11012/62154?show=full (accessed on 26 February 2021).
29. Sabitha, S.; Rajasree, M.S. Anonymous-CPABE: Privacy Preserved Content Disclosure for Data Sharing in Cloud. In *Architecture of Computing Systems—ARCS 2015*; Pinho, L.M.P., Karl, W., Cohen, A., Brinkschulte, U., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 146–157.
30. Marin, E.; Mustafa, M.A.; Singelée, D.; Preneel, B. A Privacy-Preserving Remote Healthcare System Offering End-to-End Security. In *Ad-hoc, Mobile, and Wireless Networks*; Mitton, N., Loscri, V., Mouradian, A., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 237–250.
31. Danezis, G.; Goldberg, I. Sphinx: A Compact and Provably Secure Mix Format. In Proceedings of the 2009 30th IEEE Symposium on Security and Privacy, Oakland, CA, USA, 17–20 May 2009; pp. 269–282. [CrossRef]
32. Baek, S.; Seo, S.H.; Kim, S. Preserving Patient's Anonymity for Mobile Healthcare System in IoT Environment. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 2171642. [CrossRef]
33. Donet Donet, J.A.; Pérez-Solà, C.; Herrera-Joancomartí, J. The Bitcoin P2P Network. In *Financial Cryptography and Data Security*; Böhme, R., Brenner, M., Moore, T., Smith, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 87–102.
34. Sun, Y.; Edmundson, A.; Vanbever, L.; Li, O.; Rexford, J.; Chiang, M.; Mittal, P. RAPTOR: Routing attacks on privacy in tor. In *24th USENIX Security Symposium (USENIX Security 15)*, USENIX: Washington, DC, USA, 2015, pp. 271–286.
35. Wood, G. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Proj. Yellow Paper* **2014**, *151*, 1–32.
36. Gencer, A.E.; Basu, S.; Eyal, I.; Van Renesse, R.; Sirer, E.G. Decentralization in bitcoin and ethereum networks. In *International Conference on Financial Cryptography and Data Security*; Springer: Berlin/Heidelberg, Germany, 2018, pp. 439–457.
37. Benhamouda, F.; Halevi, S.; Halevi, T. Supporting private data on hyperledger fabric with secure multiparty computation. *Ibm J. Res. Dev.* **2019**, *63*, 3:1–3:8. [CrossRef]
38. Tanwar, S.; Parekh, K.; Evans, R. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *J. Inf. Secur. Appl.* **2020**, *50*, 102407. [CrossRef]
39. Attia, O.; Khoufi, I.; Laouiti, A.; Adjih, C. An IoT-Blockchain Architecture Based on Hyperledger Framework for Healthcare Monitoring Application. In Proceedings of the 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Canary Islands, Spain, 24–26 June 2019; pp. 1–5.
40. Moubarak, J.; Filiol, E.; Chamoun, M. On blockchain security and relevant attacks. In Proceedings of the 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), Jounieh, Lebanon, 18–20 April 2018; pp. 1–6.
41. Andola, N.; Gahlot, R.; Gogoi, M.; Venkatesan, S.; Verma, S. Vulnerabilities on Hyperledger Fabric. *Pervasive Mob. Comput.* **2019**, *59*, 101050. [CrossRef]
42. Kizza, J.M. Standardization and Security Criteria: Security Evaluation of Computer Products. In *Guide to Computer Network Security*; Springer International Publishing: Cham, Switzerland, 2020; pp. 351–365.
43. PARS3 Solutions GRIIS, 2020. Available online:https://griis.ca/en/solutions/pars3/ (accessed on 26 February 2021).
44. Dahl, L.T.; Katz, A.; McGrail, K.; Diverty, B.; Ethier, J.-F.; Gavin, F.; McDonald, J.T.; Paprica, P.A.; Schull, M.; Walker, J.D.; et al. The SPOR-Canadian Data Platform: A national initiative to facilitate data rich multi-jurisdictional research. *Int. J. Popul. Data Sci.* **2020**, *5*. [CrossRef]
45. ReflexD Solutions GRIIS, 2020. Available online: https://griis.ca/en/solutions/reflexd/ (accessed on 26 February 2021).
46. Ethier, J.F.; McGilchrist, M.; Barton, A.; Cloutier, A.M.; Curcin, V.; Delaney, B.C.; Burgun, A. The TRANSFoRm project: Experience and lessons learned regarding functional and interoperability requirements to support primary care. *Learn. Health Syst.* **2018**, *2*, e10037. [CrossRef] [PubMed]
47. Lodderstedt, T.; McGloin, M.; Hunt, P. OAuth 2.0 Threat Model and Security Considerations. RFC 6819, RFC Editor, 2013. Available online: https://tools.ietf.org/html/rfc6819 (accessed on 26 February 2021).